



Conference Article

Comparative Analysis of Baseline Vnet and Unet Architectures on Pancreas Segmentation

Azim Uslucuk^{1*}, Hakan Öcal²

¹ Intelligent Systems Engineering,
Graduate School, Bartın University, Bartın, Türkiye Orcid ID: <https://orcid.org/0009-0000-0724-3337>,
e-mail: 22619824014@ogrenci.bartın.edu.tr,

² Computer Engineering, Faculty of Engineering Architecture, and Design, Bartın University, Bartın,
Türkiye, Orcid ID: <https://orcid.org/0000-0002-8061-8059>, e-mail: hocal@bartın.edu.tr,

* Correspondence: Azim Uslucuk; e-mail: 22619824014@ogrenci.bartın.edu.tr; Phone: +905423581256

(First received October 22, 2023 and in final form December 24 2023)

3rd International Conference on Design, Research and Development

(RDCONF 2023)

December 13 - 15, 2023

Reference: Uslucuk, A., Öcal, H. Comparative Analysis of Baseline Vnet and Unet Architectures on Pancreas Segmentation. *Orclever Proceedings of Research and Development*,3(1), 146-157.

Abstract

The pancreas is one of the vital organs in the human body. It has an essential role in the digestive system and endocrine system. Diseases such as cancer, diabetes, hormonal problems, pancreatitis, and digestive problems occur in pancreatic disorders. In detecting pancreatic disorders, first blood and urine tests are requested. If further examination is needed, CT (Computed Tomography), MR (Magnetic Resonance), and EUS (Endoscopic Ultrasonography) imaging methods are used. Pancreas segmentation is generally the process of defining and drawing the lines of the pancreas from medical images such as CT and MRI. The size and shape of the pancreas varies from person to person. Manual segmentation of the pancreas is time-consuming and varies between physicians. Recently, deep learning-based segmentation methods that achieve high-performance results in organ segmentation have become trendy. In this study, Unet and Vnet architectures were comparatively analyzed on the NIH-CT-82 dataset. As a result of the ablation studies, a validation sensitivity of 0.9978 and a validation loss of 0.041 were obtained in the Unet architecture. In the training with the Vnet architecture, 0.9975 validation sensitivity and 0.046 validation loss values were obtained, respectively.



Keywords: Artificial Intelligence, Deep Learning, Artificial Neural Networks, Pancreas, Segmentation.

1. Introduction

CT images are the first medical imaging used to diagnose pancreatic disorders. CT images are taken quickly and, in some cases, allow images with higher contrast to be obtained than MRI. Medical imaging is a non-invasive technique for examining internal organs and is the most common form of examination used after laboratory tests [1], [2]. The pancreas is one of the organs whose boundaries are difficult to determine due to its irregular shape and dimensions that vary from person to person [3]. Accurate segmentation of the pancreas, which occupies a small portion of CT images (<0.5%), is critical for the diagnosis and treatment planning of Pancreatic Cancer, a highly fatal disease [4], [5]. In the segmentation process of medical images, manual segmentation of large amounts of images has been extensively studied by medical image reporting groups due to concerns that it may cause distraction [6]. In order to prevent this situation, it is essential to turn to computation-based systems. Although the concept of artificial intelligence, first put forward in the 1950s, entered a winter period from time to time, its development continued to accelerate after 2000. Deep learning, one of AI's sub-branches, is a prominent method in medical image processing. Studies in the field of image analysis have shown that the success rates of segmentation with deep learning are high [7]. Computer systems that imitate the human brain's learning style and processing logic are called ANNs (artificial neural networks). Some ANNs used in pancreas segmentation are as follows. CNNs (Convolutional neural networks) are used to find local features in images. It generally defines the edges, textures, and shapes of images. CNN-based ANNs have been proposed many times in the processing of medical images [8]. Densely Connected Neural Networks(DNN) are used to process high-level features of images. U-net is a network that generally combines the features of CNN and DNN [9]. Milletari et al. proposed the Vnet network, which is a more flexible network for learning how to process 3D images volumetrically [10]. In this study, Vnet and Unet architectures, which are most used in segmentation, were comparatively analyzed.

In recent years, deep learning models consisting of convolutional neural network layers have shown high segmentation performance [11]. The deep learning method, which labels and makes inferences by identifying pixels of images of organs or lesions in medical images, is significant in terms of its high success rate [12]. Derin et al. They comparatively tested U-net and its different versions, Attention U-Net, Residual U-Net,



Attention Residual U-Net, and Residual U-Net++, using CT images from the NIH-CT82 dataset of 82 patients. They reported that the results of Residual U-Net stand out with the highest score of 0.908 precision and 0.999 accuracy [3]. Wang et al. They proposed a dual-input v-mesh fully convolutional network (FCN) to cope with the low texture contrast that makes the segmentation task difficult. They reported that this method is more performant than previous methods [4]. Paithane and Kakarwal used 12-layer deep learning networks with four convolution layers in the LMNS-Net model for pancreas segmentation and obtained a membrane similarity index score of $88.68 \pm 57.49\%$ [13]. In a new deep learning method proposed for gross tumor volume segmentation from MRI images (GTV) of pancreatic cancer patients, 126 image sets of 21 patients were used as the data set. SegResNet, SegResNet 2D, and SwinUNETR were compared as DL architectures. As a result, it was stated that training DICE = 0.88 and test DICE = 0.78 scores were obtained with the SwinUNETR model [14]. To reveal bottlenecks in pancreas segmentation with deep learning, Zhang et al. Their review of ~51 articles examined guidance and collaboration incentives to overcome the challenges of pancreas segmentation algorithms [15].

2. Materials and Methods

This section will give comprehensive information about the data set and deep learning architectures used for comparative analysis.

2.1. Preparing the dataset

The NIH-CT82 dataset consists of 82 3D CT abdominal contrast images obtained by the National Institutes of Health Clinical Center from 53 male and 27 female subjects [16]. While 17 of the subjects have healthy kidneys, 65 of them have healthy pancreas. The ages of the cases ranged from 18 to 76 years. CT images are 512x512 pixels in size. First, the data set is split into 80% training and 20% validation. Various sizing and image processing techniques were applied to the data set. Besides, images with the nifti format were converted to .npy data formats to address computational limitations.



Additionally, CT images were resized to 160x160x160. During resizing, the ROI region was cropped and enlarged. A random flip along the axis was applied to the images. Sample images from the dataset are shown in Figure 1.

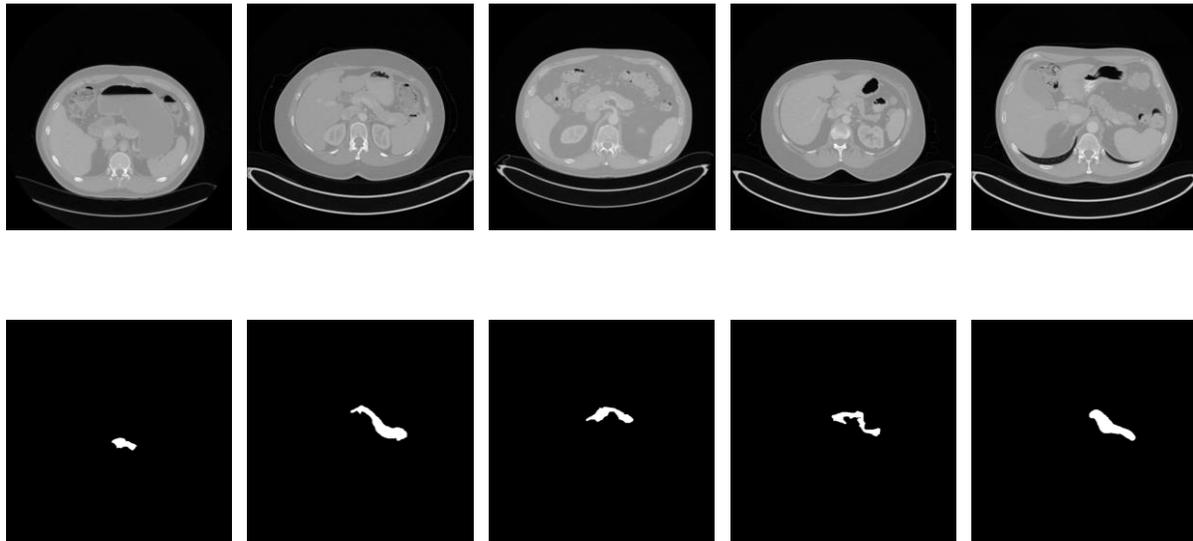


Figure 1. Sample images from the dataset

2.2 Details of the Vnet and Unet Architectures

The Baseline Unet architecture used for comparative analysis is shown in Figure 2. As seen in Figure 2, a Unet architecture ranging from 16 to 256 filters was used to segment pancreas images. ReLU was used as a batch normalization and activation function in both architectures to normalize the data. In the Unet architecture, 2x2x2 MaxPooling and dropout value was used as 0.8. Double convolution was used in each layer of Unet. In addition, in the Unet and Vnet architectures, 5x5x5 filters were used in the first and last layers, while 2x2x2 convolutional filters were used in the other layers. The Vnet architecture used in the study is shown in Figure 3. As seen in Figure 3, a Vnet architecture ranging from 16 to 256 filters was used to segment pancreas images. Vnet architecture, unlike Unet architecture, processes 3D images volumetrically. In addition, Vnet differs from U-Net in that it uses the convolutions layer instead of the up-sampling and down-sampling pooling layer. The idea behind V-Net is that using the Maxpooling process causes much information loss, so replacing it with another series of convolution operations without padding helps preserve more information.

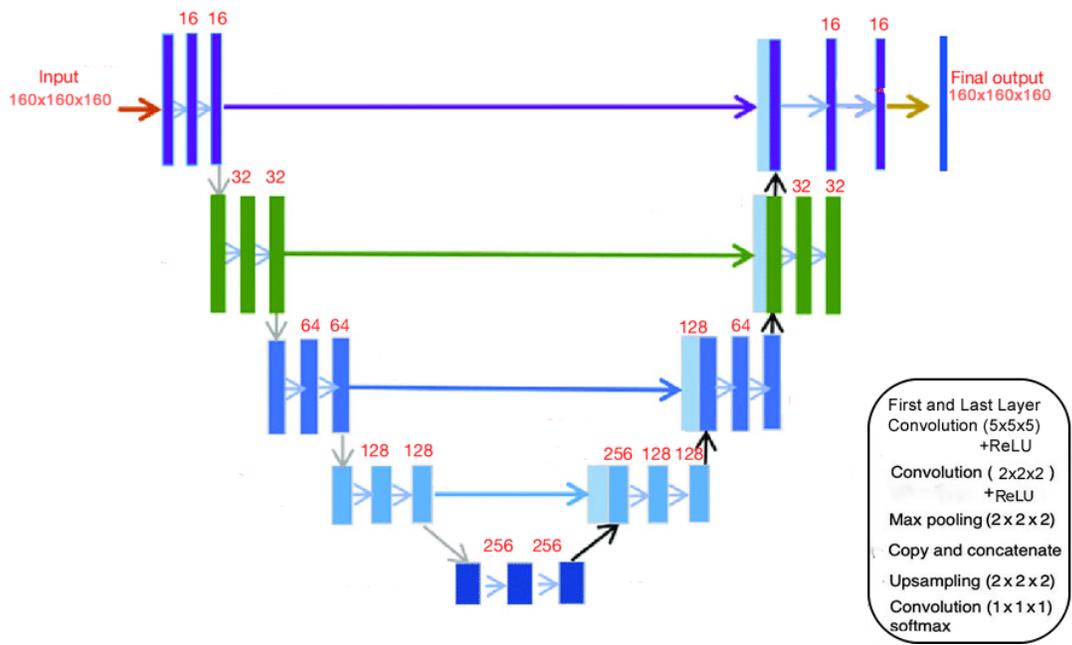


Figure 2. Baseline U-net Architecture

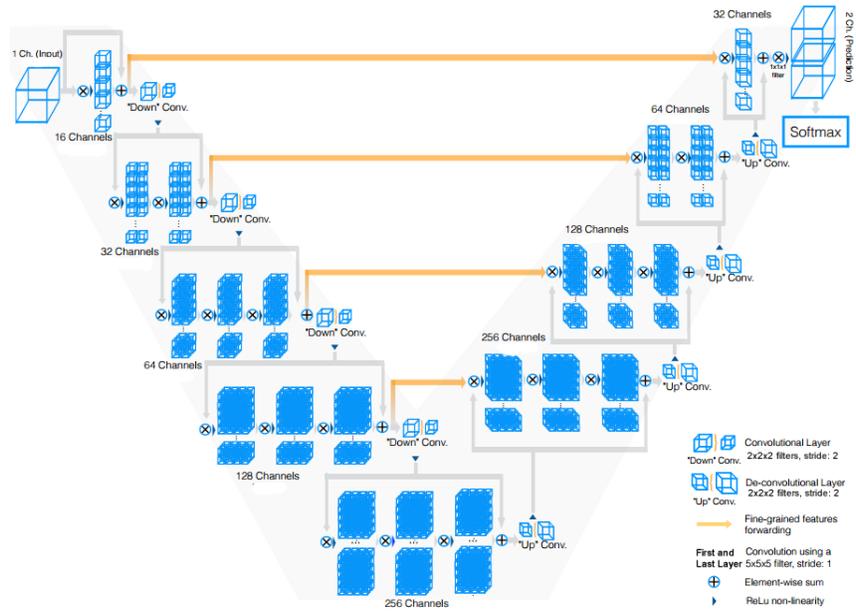


Figure 3. Baseline V-net Architecture



2.4 Performance metrics

Accuracy metrics were used for comparative performance evaluations of the models. The mathematical formulas of the metrics are shown in Equation 1. In the equations, FP means False Positive, FN means False Negative, TP means True Positive, and TN means True Negative.

$$ACC = \frac{TN+TP}{TN+TP+FN+FP} \quad (1)$$

3. Result

This section gives comparative performance analyses of the employed methodologies and comparative information about the models' good and bad aspects.

3.1 Model's implementation details

Training and validation of the recruited models were carried out on the NVidia RTX 4000 graphics card. For the models, batch_size is one, optimizer ADAM, and the learning rate is 0.001 [17]. The proposed models were trained on the dataset for 250 epochs. The algorithms of the models were implemented in the Anaconda ecosystem with the Python 3.8 programming language and Tensorflow_gpu 2.5 library. Sparse Categorical Cross entropy loss was used to calculate the loss of the models.

3.2 Ablation studies

Considering the computational limitations in the inference studies, the models were designed as U and V-shaped with 16 to 256 filters. It was observed that the training failed when the model size was increased by further reducing the data set size. In the studies conducted, dropout values for Unet were determined as 0.5 to 0.9, and 0.8 was selected as the ideal value in the training. For Vnet, dropout values from 0.1 to 0.9 were applied, and 0.2 was determined to be the perfect dropout value.



3.3 Comparative performance analysis on the NIH-CT82 dataset

Figure 4 and Figure 5 show the training accuracy and training loss of the Baseline Unet architecture. The training sensitivity of Baseline Unet varied between 99.78 and 99.76. The educational loss of the hired architect started from 0.18 and decreased to 0.05. No decrease in the training loss of the architecture was observed after 250 epochs.

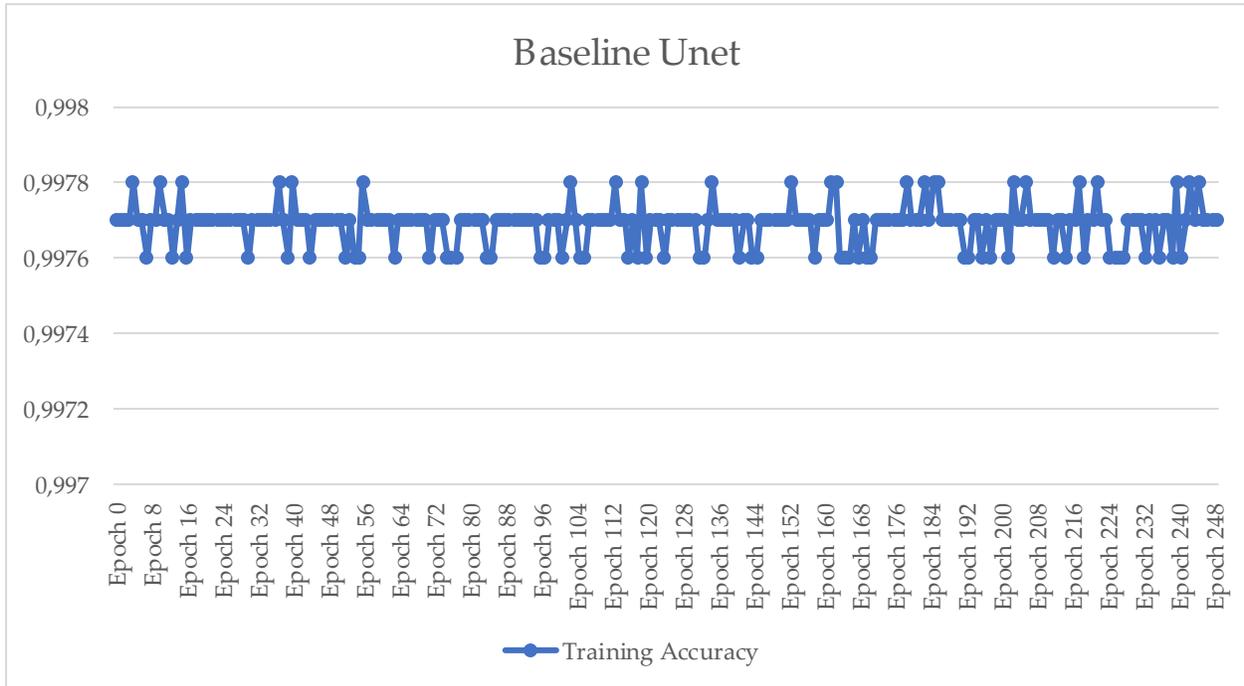


Fig. 4. Baseline Unet Training Accuracy NHI-CT82 dataset

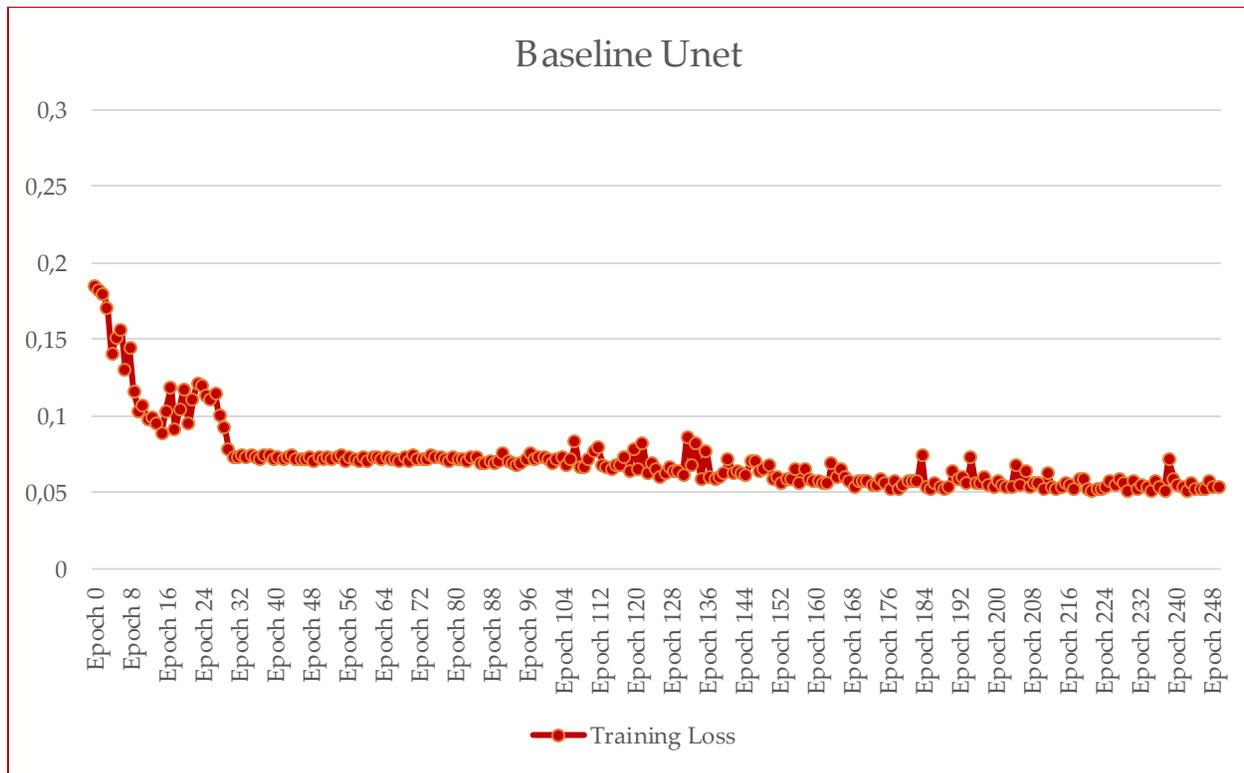


Fig. 5. Baseline Unet Training Loss NHI-CT82 dataset

Baseline Vnet architecture performed slightly better than Unet architecture despite the high parameter values it used. Figure 6 and Figure 7 show the training accuracy and training loss of the Baseline Vnet architecture, respectively. The designed model achieved a performance value of 99.60 in training accuracy and a performance value of 0.036 in training loss. One of the biggest reasons the Vnet architecture uses high parameters is the convolution process instead of the Maxpooling layer used in the Unet architecture.

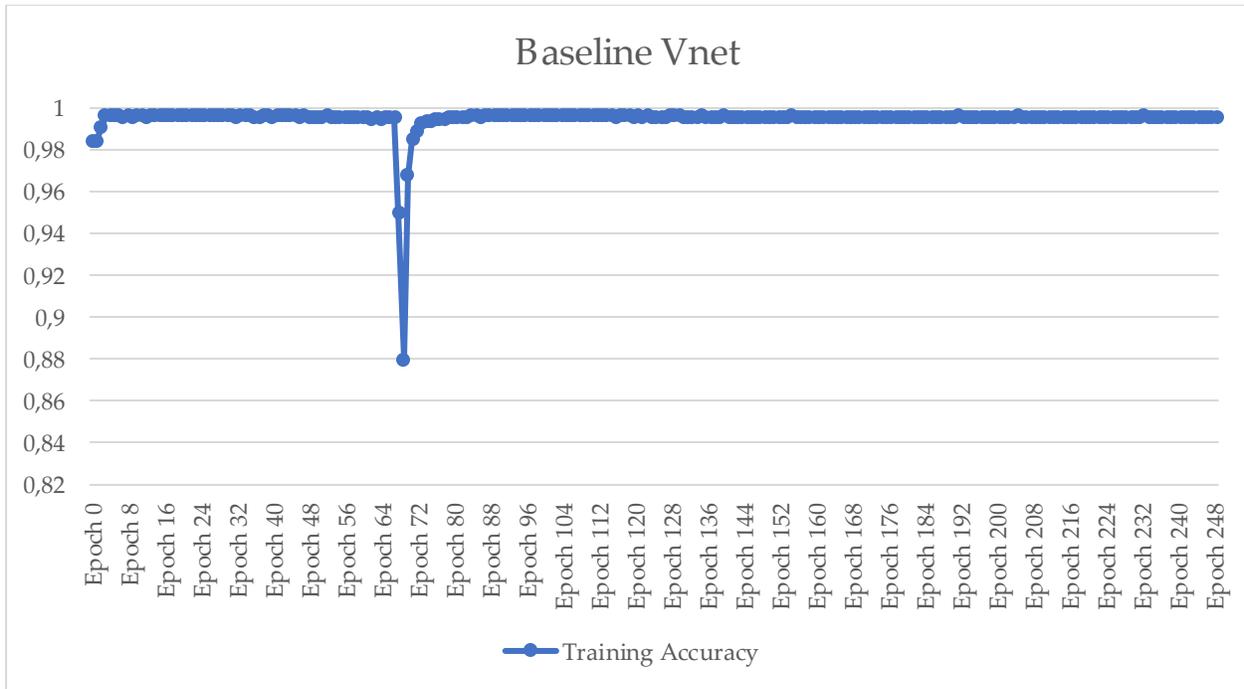


Fig. 6. Baseline Vnet Training Accuracy on NHI-CT82 dataset

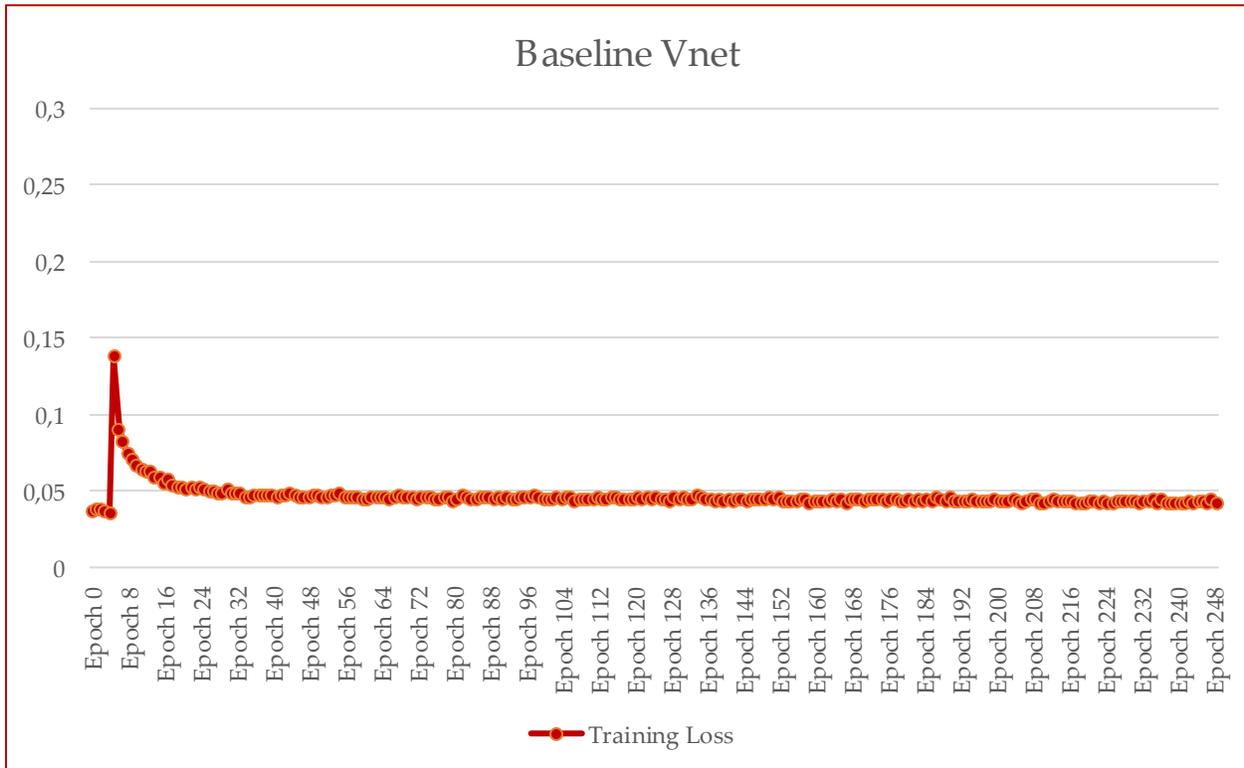


Fig. 7. Baseline Vnet Training Loss on NHI-CT82 dataset



Comparative analysis results of Unet and Vnet architectures are shown in Table 1. As a result of the analysis, Vnet, which has 22.1 million parameters, performed slightly better than Unet, which has 2.94 million parameters. Due to the number of parameters, training Vnet took almost four times longer.

Table 1. Comparative performance results of the models

Architectures	Methods	Parameters(M)	Training Accuracy(%)	Validation Accuracy(%)	Epoch
Baseline Unet	CT images	2.94	99.78	99.78	250
Baseline Vnet	CT images	22.1	99.6	99.82	

4. Discussion and Conclusion

In this study, Unet and Vnet architectures, which are most commonly used in the literature with their various variations, were comparatively analyzed on NHI-CT82 Pancreas CT images. The analysis also showed that both models were robust in terms of segmentation. However, as can be seen from the comparative analysis, Unet performed the segmentation process with 7.5 times fewer parameters than Vnet. Additionally, training for Vnet’s architecture took 4 four times longer than training for Unet’s architecture. One of the main reasons for this situation is that the Vnet architecture, unlike the Unet architecture, uses the Convolution process instead of the Maxpooling layer. The main reason for using a convolution layer instead of Maxpooling in the Vnet architecture is that Maxpooling may cause high-level features to be lost. However, as can be seen from the performance results, these two types of architecture continue to be indispensable in all kinds of segmentation tasks, especially the segmentation of medical images.



References

- [1] H. Kasban, M. A. M. El-Bendary, and D. H. Salama, "A Comparative Study of Medical Imaging Techniques," *Int. J. Inf. Sci. Intell. Syst.*, vol. 4(2), pp. 37–58, 2015, [Online]. Available: https://ilearn.th-deg.de/pluginfile.php/480243/mod_book/chapter/8248/updated_JXISIS2015.pdf
- [2] M. Aljabri and M. AlGhamdi, "A review on the use of deep learning for medical images segmentation," *Neurocomputing*, vol. 506, pp. 311–335, Sep. 2022, doi: 10.1016/j.neucom.2022.07.070.
- [3] A. Derin, C. Gurkan, A. Budak, and H. KARATAŞ, "Pancreas Segmentation Using U-Net Based Segmentation Networks in CT Modality: A Comparative Analysis," *Avrupa Bilim Ve Teknol. Derg.*, no. 40, pp. 94–98, 2022, Accessed: Oct. 07, 2023. [Online]. Available: <https://dergipark.org.tr/en/pub/ejosat/article/1171803>
- [4] Y. Wang *et al.*, "Pancreas segmentation using a dual-input v-mesh network," *Med. Image Anal.*, vol. 69, p. 101958, Apr. 2021, doi: 10.1016/j.media.2021.101958.
- [5] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. K. Fishman, and A. L. Yuille, "A Fixed-Point Model for Pancreas Segmentation in Abdominal CT Scans," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, Eds., in *Bilgisayar Bilimleri Ders Notları*. Cham: Springer International Publishing, 2017, pp. 693–701. doi: 10.1007/978-3-319-66182-7_79.
- [6] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas." arXiv, May 20, 2018. Accessed: Oct. 26, 2023. [Online]. Available: <http://arxiv.org/abs/1804.03999>
- [7] T. ŞENTÜRK and F. LATİFOĞLU, "Biyomedikal Görüntülerin Bölütlenmesine Yönelik Derin Öğrenmeye Dayalı Yöntemler: Bir Gözden Geçirme," *Dicle Üniversitesi Fen Bilim. Enstitüsü Derg.*, vol. 12, no. 1, pp. 161–187, 2023, Accessed: Oct. 07, 2023. [Online]. Available: <https://dergipark.org.tr/en/pub/dufed/issue/75551/1181996>
- [8] F. Yuan, Z. Zhang, and Z. Fang, "An effective CNN and Transformer complementary network for medical image segmentation," *Pattern Recognit.*, vol. 136, p. 109228, Apr. 2023, doi: 10.1016/j.patcog.2022.109228.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in *Lecture Notes in Computer Science*, vol. 9351., Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
- [10] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, Oct. 2016, pp. 565–571. doi: 10.1109/3DV.2016.79.
- [11] Ayhan M S, Kühlewein L, Aliyeva G, Inhoffen W, Ziemssen F, Berens P. "Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection". *Med. Image Anal.* 64, 101724, 2020.



- [12]A. G. Eker and N. Duru, "Medikal Görüntü İşlemede Derin Öğrenme Uygulamaları," *Acta Infologica*, vol. 5, no. 2, Art. no. 2, Dec. 2021, doi: 10.26650/acin.927561.
- [13]P. Paithane and S. Kakarwal, "LMNS-Net: Lightweight Multiscale Novel Semantic-Net deep learning approach used for automatic pancreas image segmentation in CT scan images," *Expert Syst. Appl.*, vol. 234, p. 121064, Dec. 2023, doi: 10.1016/j.eswa.2023.121064.
- [14]W. Choi *et al.*, "Novel Deep Learning Segmentation Models for Accurate GTV and OAR Segmentation in MR-Guided Adaptive Radiotherapy for Pancreatic Cancer Patients," *Int. J. Radiat. Oncol.*, vol. 117, no. 2, Supplement, p. e462, Oct. 2023, doi: 10.1016/j.ijrobp.2023.06.1660.
- [15]Z. Zhang, L. Yao, E. Keles, Y. Velichko, and U. Bagci, "Deep Learning Algorithms for Pancreas Segmentation from Radiology Scans: A Review," *Adv. Clin. Radiol.*, vol. 5, no. 1, pp. 31–52, Sep. 2023, doi: 10.1016/j.yacr.2023.05.001.
- [16]H. R. Roth *et al.*, "DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation." arXiv, Jun. 21, 2015. Accessed: Dec. 05, 2023. [Online]. Available: <http://arxiv.org/abs/1506.06448>
- [17]D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR*, Dec. 2014, Accessed: Dec. 05, 2023. [Online]. Available: <https://www.semanticscholar.org/paper/a6cb366736791bcccc5c8639de5a8f9636bf87e8>